SECTION I

ISOLATED RABBIT EYE (IRE) TEST METHOD

ACCURACY AND RELIABILITY REANALYSIS

[This Page Intentionally Left Blank]

1.0 INTRODUCTION

On November 1, 2004, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) released draft Background Review Documents (BRDs) on the current status of four *in vitro* test methods for detecting ocular corrosives and severe irritants (see http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm). The test methods reviewed were the Bovine Corneal Opacity and Permeability (BCOP), the Hen's Egg Test - Chorioallantoic Membrane (HET-CAM), the Isolated Rabbit Eye (IRE), and the Isolated Chicken Eye (ICE) assays. On January 11-12, 2005, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) convened an Expert Panel to independently evaluate the validation status of these four *in vitro* test methods for identifying ocular corrosives or severe irritants. The Expert Panel Report, *Evaluation of the Current Validation Status of In Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants*, can be obtained directly from NICEATM or electronically from http://iccvam.niehs.nih.gov/methods/eyeirrit.htm). Public comments at the meeting revealed that additional data could be made available that had not yet been provided in response to earlier requests for data. The Expert Panel subsequently recommended that the additional data be requested and that a reanalysis of the accuracy and reliability of each test method be conducted, to the extent possible.

In response to this recommendation, a second *Federal Register* (*FR*) notice was published on February 28, 2005 (*FR* Vol. 70, No. 38, pp. 9661-9662; http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) requesting all available *in vitro* data on these four *in vitro* ocular irritancy test methods and corresponding *in vivo* rabbit eye test method data, as well as any human exposure data (either via ethical human studies or accidental exposure). The first *FR* notice requesting these data had been published on March 24, 2004 (*FR* Vol. 69, No. 57, pp. 13859-13861; http://iccvam.niehs.nih.gov/methods/eyeirrit.htm). Also, a request for relevant data was re-sent directly to the primary developers or users of each test method and sent to other scientists who participated in or attended the Expert Panel Meeting on January 11-12, 2005 and who had indicated a desire to provide additional data. No human exposure data was obtained for the substances evaluated in the IRE test method, and therefore no calculations could be made for the accuracy of the IRE test method for predicting human severe ocular irritancy.

Other factors also necessitated a reanalysis of the accuracy of the IRE test method for detecting ocular corrosives and severe irritants. First, clarification regarding the rules for classification of severe irritants was obtained subsequent to the release of the four BRDs that resulted in changes to the hazard classification of some of the substances used in the original analysis. For the original analysis, reversibility of ocular effects for the European Union (EU) and United Nations (UN) Globally Harmonized System (GHS) hazard classification systems was considered to be achieved if, by post-exposure day 21, the endpoint scores fell below the threshold that resulted in a test substance being classified as a severe irritant (EU [2001]; UN [2003]). The new information obtained indicated that reversibility of ocular effects is achieved only when all scores reach zero by post-exposure day 21. This change

resulted in nine substances previously classified as EU nonsevere irritants now being classified as EU severe irritants. One substance previously classified as GHS nonsevere irritant was reclassified as GHS severe irritant.

Second, the chemical classes assigned to each test substance were revised to reflect a standardized classification scheme (based on the Medical Subject Headings [MeSH]; http://www.nlm.nih.gov/mesh) that would ensure consistency in classifying substances among all *in vitro* ocular test methods under consideration. This resulted in some chemicals being re-classified. The accuracy of the IRE test method, by chemical class and using the GHS classification system (UN [2003]), has been reanalyzed to reflect these changes.

Finally, an additional accuracy analysis was conducted. In this analysis, the accuracy of each *in vitro* ocular irritancy test method for detecting ocular corrosives or severe irritants, depending on whether the classification was based on the severity of the response and/or its persistence to day 21 post-exposure, was determined.

For the IRE test method, the changes to the existing database that resulted from using the appropriate persistence classification criteria and any new data and/or information received subsequent to the release of the draft BRD are summarized in **Table I-1**. For the IRE test method, the changes to the existing database that resulted from using the appropriate persistence classification criteria and any new information received in response to the Expert Panel meeting and to additional requests for information are summarized in **Table I-1**.

No additional comparative *in vitro-in vivo* test results data were submitted for the IRE test method. The existing database of substances tested using the four ocular endpoints recommended in the draft IRE BRD (corneal opacity, corneal swelling, fluorescein penetration, and epithelial integrity) remained limited to the Guerriero et al. (2004) data set. However, as recommended by the Expert Panel, a reanalysis was performed in which substances in the CEC (1991), Balls et al. (1995), and Gettings et al. (1996) studies that had been identified as corrosives/severe irritants using appropriate decision criteria (a corneal opacity score greater than or equal to 3, or a corneal swelling equal to or greater than a 25%) were considered together with the test results obtained by Guerriero et al. (2004). This database is referred to as the "Expanded Data Set."

Substances that were identified as corrosives/severe irritants based on *in vitro* results by any single endpoint were, therefore, included in the reanalysis as part of the "Expanded Data Set." Substances in CEC (1991), Balls et al. (1995), and Gettings et al. (1996) that were identified as nonsevere irritants, based on *in vitro* results, were not included in the "Expanded Data Set," because any of the omitted endpoints might have resulted in a severe irritant classification. For example, in Gettings et al. (1996), only corneal swelling was measured. Substances that produced corneal swelling ≥ 25% were classified as severe irritants and were included in the "expanded data set." However, a substance that did not produce ≥ 25% corneal swelling, might have produced a corneal opacity score, fluorescein penetration score, or damage of the epithelium that would have classified it as a severe irritant had those endpoints been evaluated.

92 **Table I-1. Summary of IRE Database Changes**
93

Data Source	Data Set	Number of Available Substances	Number of Acceptable Substances by Ocular Irritancy Classification System			Comments
			EPA[1]	EU[2]	GHS[3]	
			Cat[4] I/Total[5]	R41/Total	Cat 1/Total	
CEC (1991)[6]	New[7]	21	-	5/15	-	Six substances were excluded from the original database (n=21) because their EU classification was based on pH extreme or skin corrosivity information rather than *in vivo* rabbit eye test data.
	Old[7]	21	-	11/21	-	
Balls et al. (1995)	New	59	19/53	19/49	22/54	The decrease in the total number of usable substances is due to excluding substances from consideration due to insufficient rabbit eye test data for classification (See **Appendix I-A**).
	Old	59	20/54	21/59	22/56	
Gettings et al. (1996)	New	25	17/25	16/24	16/24	The increase in the number of corrosive/severe irritants is due to the reclassification of several substances based on the presence of ocular damage at day 21 post-treatment.
	Old	25	12/25	12/25	12/25	
Guerriero et al. (2004)	New	44	11/38	11/38	11/38	Six substances were excluded from the original database because their classification was based on pH extremes or skin corrosivity information rather than *in vivo* rabbit eye test data.
	Old	44	16/41	15/41	16/41	
Expanded Data Set[8]	New	91[1]	31/76	37/80	33/76	From 11-15 substances were excluded from the original database, because specific regulatory classification criteria were not met (e.g., persistence could not be determined due to study termination).

94 [1]EPA = U.S. Environmental Protection Agency (EPA [1996]).
95 [2]EU = European Union (EU [2001]).
96 [3]GHS = Globally Harmonized System (UN [2003]).
97 [4]Cat = Category.
98 [5]Number of severe irritants by regulatory classification/number of classifiable substances.
99 [6]When the same substance was evaluated in multiple laboratories, the IRE ocular irritancy potential for each independent test result was determined.
100 Subsequently, an overall IRE ocular irritancy classification was assigned for each substance based on the majority of ocular irritancy classification calls and this
101 call was used in the analysis of IRE test method accuracy (approach described in **Section I-2.1**).
102 [7]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft IRE BRD.

[8]Includes the 38 substances tested by Guerriero et al. (2004) that could be classified and additional substances classified as severe irritants from CEC (1991) (EU classification system only), Balls et al. (1995), and Gettings et al. (1996), based either on an *in vitro* corneal opacity score of at least 3.0 or an *in vitro* corneal swelling of at least 25%; these were among the criteria used by Guerriero et al. (2004) to identify corrosive/severe irritants.

2.0 ACCURACY OF THE IRE TEST METHOD - REANALYSIS

The ability of the IRE test method to correctly identify ocular corrosives and severe irritants, as defined by the U.S. Environmental Protection Agency (EPA), EU, and GHS classification systems (EPA [1996]; EU [2001]; UN [2003])[1], was evaluated. The three regulatory ocular hazard classification systems considered during this analysis use different classification systems and decision criteria to identify ocular corrosives and severe irritants based on *in vivo* rabbit eye test results. All three classification systems are based on individual animal data in terms of the magnitude of the response and on the extent to which induced ocular lesions fail to reverse by day 21. However, there are differences among the three classification systems in regard to their criteria used by NICEATM for distinguishing between a severe and a nonsevere response (see **Appendix A**). Thus, to evaluate the accuracy of the IRE test method for identifying ocular corrosives and severe irritants, individual rabbit data collected at the different observation times was needed for each substance.

The ability of the IRE test method to correctly identify ocular corrosives and severe irritants, as defined by the EPA, EU, and GHS classification systems (EPA [1996]; EU [2001]; UN [2003]), was evaluated using two approaches. In the first approach, the accuracy of IRE was assessed separately for each *in vitro-in vivo* comparative study (i.e., publication) reviewed in Sections 4.0 and 5.0 of the draft IRE BRD. In the second approach, an overall analysis of IRE test method accuracy was conducted by combining results from each study, and then an overall ocular irritancy classification was assigned for each substance. When the same substance was evaluated in multiple laboratories, the overall IRE ocular irritancy classification was based on the majority of calls among all of the studies. When there was an even number of different irritancy classifications for substances (e.g., two tests classified a substance as a nonsevere irritant and two tests classified a substance as a severe irritant), the more severe irritancy classification was used for the overall classification for the substance (severe irritant, in this case).

Based on the revisions made to the IRE and *in vivo* test method databases, a revised accuracy analysis has been conducted. The calculations were performed as described previously in Section 6.0 of the draft IRE BRD. To allow for a comparison of the results obtained in the revised analysis relative to those obtained previously, the data tables below include accuracy statistics from both analyses. However, the discussion of the results in the sections that follow relate to the revised analysis only.

2.1 GHS Ocular Hazard Classification System

Three studies (Balls et al. [1995]; Gettings et al. [1996]; Guerriero et al. [2004]) contained IRE test data on 128 substances, 116 of which had sufficient *in vivo* data to be assigned an

[1] For the purposes of this analysis, an ocular corrosive or severe irritant was defined as a substance that would be classified as Category 1 according to the GHS classification system (UN [2003]), as Category I according to the EPA classification system (EPA [1996]), or as R41 according to the EU classification system (EU [2001]).

ocular irritancy classification as defined by the GHS classification system (UN [2003])[2] (see **Appendix I-A**). Based on results from *in vivo* rabbit eye experiments, 49[3] of the 116 substances were classified as severe irritants (i.e., Category 1), the other 67 substances were classified as nonsevere irritants (either Category 2A, 2B) or nonirritants (**Table I-2**). The 12 substances that could not be classified according to the GHS classification system due to the lack of adequate animal data are so noted in **Appendix I-A**.

2.1.1 Balls et al. (1995)

Based on the reclassification process, 54 of the 59 substances tested in this study could be assigned a GHS classification (**Table I-2**). The remaining five substances had inadequate *in vivo* data for assigning a classification according to the GHS system (UN [2003]). For the 54 substances assigned a GHS classification, the IRE test method has an accuracy[4] of 54% (29/54), a sensitivity of 68% (15/22), a specificity of 44% (14/32), a false positive rate of 56% (18/32), and a false negative rate of 32% (7/22).

2.1.2 Gettings et al. (1996)

Based on the reclassification process, 24 of the 25 substances tested in this study could be assigned a GHS classification **(Table I-2)**. The remaining substance had inadequate *in vivo* data for assigning a classification according to the GHS system (UN [2003]). For the 24 substances that could be evaluated, the IRE test method has an accuracy of 67% (16/24), a sensitivity of 63% (10/16), a specificity of 75% (6/8), a false positive rate of 25% (2/8), and a false negative rate of 38% (6/16).

2.1.3 Guerriero et al. (2004)

Based on the reclassification process, 38 of 44 substances tested in this study could be assigned a GHS classification (**Table I-2**). The remaining six substances had inadequate *in vivo* data for assigning a classification according to the GHS system (UN [2003]). For the 38 substances that could be evaluated, the IRE test method has an accuracy of 79% (30/38), a sensitivity of 100% (11/11), a specificity of 70% (19/27), a false positive rate of 30% (8/27), and a false negative rate of 0% (0/11).

---

[2] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify GHS Category 1 irritants (i.e., severe irritants); substances classified as GHS Category 2A and 2B irritants were identified as nonsevere irritants.

[3] One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice in the same laboratory. The results were discordant with respect to GHS classification. According to one test, the classification was Category 1, while results from the other test yielded a Category 2B classification. The accuracy analysis was performed with the substance classified as Category 1.

[4] Accuracy is defined as the proportion of correct outcomes (positive and negative) of a test method; Sensitivity is defined as the proportion of all positive substances that are classified as positive; Specificity is defined as the proportion of all negative substances that are classified as negative; Positive predictivity is defined as the proportion of correct positive responses among substances testing positive; Negative predictivity is defined as the proportion of correct negative responses among substances testing negative; False positive rate is defined as the proportion of all negative substances that are falsely identified as positive; False negative rate is the defined as the proportion of all positive substances that are falsely identified as negative.

Table I-2. Evaluation of the Performance of the IRE Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS[1] Classification System, by Study and Overall

Data Source	Data Set	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al. (1995)[4,5]	New[6]	54/59	54	29/54	68	15/22	44	14/32	45	15/23	67	14/21	56	18/32	32	7/22
	Old[6]	56/59	50	28/56	64	14/22	41	14/34	41	14/34	64	14/22	59	20/34	36	8/22
Gettings et al. (1996)	New	24/25	67	16/24	63	10/16	75	6/8	83	10/12	50	6/12	25	2/8	38	6/16
	Old	25/25	64	16/25	56	9/16	78	7/9	82	9/11	50	7/14	22	2/9	44	7/16
Guerriero et al. (2004)	New	38/44	79	30/38	100	11/11	70	19/27	58	11/19	100	19/19	30	8/27	0	0/11
	Old	36/44	78	28/36	100	12/12	67	16/24	60	12/20	100	16/16	33	8/24	0	0/12
Expanded Data Set[7]	New	76/91	68	52/76	100	33/33	44	19/43	58	33/57	100	19/19	56	24/43	0	0/33

[1]GHS = United Nations Globally Harmonized System (UN [2003]).
[2]N = number of substances included in this analysis/the total number of substances in the study.
[3]Data used to calculate the percentage.
[4]One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice within the same laboratory. The results were discordant with respect to GHS classification; the analysis was performed assuming Category 1 classification.
[5]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.
[6]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on analysis included in the draft IRE BRD with corrections.
[7]Includes the 38 substances tested by Guerriero et al. (2004) that could be classified and 38 additional substances classified as severe irritants from Balls et al. (1995) and Gettings et al. (1996), based either on an *in vitro* corneal opacity score of at least 3.0 or an *in vitro* corneal swelling of at least 25%; these were among the criteria used by Guerriero et al. (2004) to identify corrosive/severe irritants. When the same substance was evaluated in multiple laboratories, the IRE ocular irritancy potential for each independent test result was determined. Subsequently, an overall IRE ocular irritancy classification was assigned for each substance based on the majority of ocular irritancy classification calls and this call was used in the analysis of IRE test method accuracy (approach described in **Section I-2.0**); this process reduced the total number of substances in the expanded data set to 76 for the GHS classification system (UN [2003]).

2.1.4 Expanded Data Set

Subsequent to the original IRE test method accuracy analysis, the total data base of 149 substances was mined to established an expanded data set that included: (1) all substances evaluated by Guerriero et al. (2004) that could be assigned an GHS classification (UN [2003]), and (ii) any additional substances classified as severe irritants by Balls et al. (1995) and Gettings et al. (1996), based either on an *in vitro* corneal opacity score of at least 3.0 or an *in vitro* corneal swelling of at least 25%, that had corresponding *in vivo* rabbit eye test data that would allow the substances to be classified according to the GHS system (UN [2003]). These two criteria were among those used by Guerriero et al. (2004) to identify corrosive/severe irritants. When the same substance was evaluated in multiple laboratories, the IRE ocular irritancy potential for each independent test result was determined. Subsequently, an overall IRE ocular irritancy classification was assigned for each substance based on the majority of ocular irritancy classification calls and this call was used in the analysis of IRE test method accuracy (approach described in **Section I-2.0**).

Using this approach, the total number of substances in the expanded data set was 76 for the GHS classification system (UN [2003]). For these 76 substances (**Table I-2**), the IRE test method has an accuracy of 68% (52/76), a sensitivity of 100% (33/33), a specificity of 44% (19/43), a false positive rate of 56% (24/43), and a false negative rate of 0% (0/33).

2.2 EPA Ocular Hazard Classification System

Three studies (Balls et al. [1995]; Gettings et al. [1996]; Guerriero et al. [2004]) contained IRE test method data on 128 substances, 116 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification according to the EPA classification system (EPA [1996])[5] (see **Appendix I-A**). Based on results from the *in vivo* rabbit eye test, 47 of these 116 substances were classified as severe irritants (i.e., Category I), while the other 69 substances were classified as nonsevere irritants or nonirritants (Categories II, III, or IV). The 12 substances that could not be classified according to the EPA classification system are so noted in **Appendix I-A**.

2.2.1 Balls et al. (1995)

Based on the reclassification process, 53 of the 59 substances tested in this study could be assigned an EPA classification (**Table I-3**). The remaining five substances had inadequate *in vivo* data for assigning a classification according to the EPA system (1996). For the 53 substances that could be evaluated, the IRE test method has an accuracy of 51% (27/53), a sensitivity of 65% (13/20), a specificity of 42% (14/33), a false positive rate of 58% (19/33), and a false negative rate of 35% (7/20).

[5] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify EPA Category I irritants (i.e., severe irritants); substances classified as EPA Category II, III, or IV irritants were defined as nonsevere irritants.

Table I-3. Evaluation of the Performance of the IRE Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA[1] Classification System, by Study and Overall

Data Source	Data Set	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al. (1995)[4,5]	New[6]	53/59	51	27/53	65	13/20	42	14/33	41	13/32	67	14/21	58	19/33	35	7/20
	Old[6]	52/59	48	25/52	61	11/18	41	14/34	35	11/31	67	14/21	59	20/34	39	7/18
Gettings et al. (1996)	New	25/25	64	16/25	59	10/17	75	6/8	83	10/12	46	6/13	25	2/8	41	7/17
	Old	25/25	60	15/25	53	9/17	75	6/8	82	9/11	43	6/14	25	2/8	47	8/17
Guerriero et al. (2004)	New	38/44	79	30/38	100	11/11	70	19/27	58	11/19	100	19/19	30	8/27	0	0/11
	Old	36/44	78	28/36	100	12/12	67	16/24	58	12/20	100	16/16	33	8/24	0	0/12
Expanded Data Set[7]	New	76/91[5]	66	50/76	100	31/31	42	19/45	54	31/57	100	19/19	58	26/45	0	0/31

[1]EPA = U.S. Environmental Protection Agency (EPA [1996]).
[2]N = number of substances included in this analysis/the total number of substances in the study.
[3]Data used to calculate the percentage.
[4]One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice within the same laboratory. The results were discordant with respect to EPA classification; the analysis was performed assuming Category I classification.
[5]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.
[6]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on analysis included in the draft IRE BRD with corrections.
[7]Includes the 38 substances tested by Guerriero et al. (2004) that could be classified and 38 additional substances classified as severe irritants from Balls et al. (1995) and Gettings et al. (1996), based either on an *in vitro* corneal opacity score of at least 3.0 or an *in vitro* corneal swelling of at least 25%; these were among the criteria used by Guerriero et al. (2004) to identify corrosive/severe irritants. When the same substance was evaluated in multiple laboratories, the IRE ocular irritancy potential for each independent test result was determined. Subsequently, an overall IRE ocular irritancy classification was assigned for each substance based on the majority of ocular irritancy classification calls and this call was used in the analysis of IRE test method accuracy (approach described in **Section I-2.0**); this process reduced the total number of substances in the expanded data set to 76 for the EPA classification system (EPA [1996]).

2.2.2 Gettings et al. (1996)

Based on the reclassification process, 25 of the 25 substances tested in this study could be assigned an EPA classification (**Table I-3**). For these 25 substances, the IRE test method has an accuracy of 64% (16/25), sensitivity of 59% (10/17), a specificity of 75% (6/8), a false positive rate of 25% (2/8), and a false negative rate of 41% (7/17).

2.2.3 Guerriero et al. (2004)

Based on the reclassification process, 38 of the 44 substances tested in this study could be assigned an EPA classification **(Table I-3)**. The remaining six substances had inadequate *in vivo* data for assigning a classification according to the EPA system (EPA [1996]). For the 38 substances that could be evaluated, the IRE test method has an accuracy of 79% (30/38), a sensitivity of 100% (11/11), a specificity of 70% (19/27), a false positive rate of 30% (8/27), and a false negative rate of 0% (0/11).

2.2.4 Expanded Data Set

Subsequent to the original IRE test method accuracy analysis, the total data base of 149 substances was mined to established an expanded data set that included: (1) all substances evaluated by Guerriero et al. (2004) that could be assigned an EPA classification (EPA [1996]), and (ii) any additional substances classified as severe irritants by Balls et al. (1995) and Gettings et al. (1996), based either on an *in vitro* corneal opacity score of at least 3.0 or an *in vitro* corneal swelling of at least 25%, that had corresponding *in vivo* rabbit eye test data that would allow the substances to be classified according to the EPA system (EPA [1996]). As noted previously, these two criteria were among those used by Guerriero et al. (2004) to identify corrosive/severe irritants. Rules for classifying a substance that was evaluated in multiple laboratories are the same as described in **Section I-2.1.4**. Based on this approach, the total number of substances in the expanded data set was 76 for the EPA classification system (EPA [1996]). For these 76 substances (**Table I-3**), the IRE test method has an accuracy of 66% (50/76), a sensitivity of 100% (31/31), a specificity of 42% (19/45), a false positive rate of 58% (26/45), and a false negative rate of 0% (0/31).

2.3 EU Ocular Hazard Classification System

Four studies (CEC [1991]; Balls et al. [1995]; Gettings et al. [1996]; Guerriero et al. [2004]) contained IRE test method data on 149 substances, 126 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification according the EU classification system (EU [2001])[6] (see **Appendix I-A**). Based on results from the *in vivo* rabbit eye test, 51[7] of the 126 substances were classified as severe irritants (i.e., R41) and the other 75 substances were classified as nonsevere irritants (either R36) or nonirritants. The two substances that could not be classified according to the EU classification system are so noted in **Appendix I-A**.

[6] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify R41 irritants (i.e., severe irritants); substances classified as R36 were defined as nonsevere irritants.

[7] One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice in the same laboratory. The results were discordant with respect to EU classification. According to one test, the classification was R41, while results from the other test yielded an R36 classification. The accuracy analysis was performed with the substance classified as R41.

2.3.1 CEC (1991)

Based on the reclassification process, 15 of the 21 substances tested in this study were included in an analysis of accuracy (**Table I-4**). The remaining six substances had inadequate *in vivo* data for assigning a classification according to the EU system (EU [2001]). Based on the available *in vivo* rabbit eye data or the EU ocular irritancy classification for each substance provided in the published study (individual rabbit eye test data was not available for all of the substances), the IRE test method has an accuracy of 87% (13/15), a sensitivity of 100% (5/5), a specificity of 80% (8/10), a false positive rate of 20% (2/10), and a false negative rate of 0% (0/5).

2.3.2 Balls et al. (1995)

Based on the reclassification process, 49 of the 59 substances tested in this study could be assigned a EU classification (**Table I-4**). The remaining ten substances had inadequate *in vivo* data for assigning a classification according to the EU system (EU [2001]). For the 49 substances assigned an EU classification, the IRE test method has an accuracy of 55% (27/49), sensitivity of 74% (14/19), a specificity of 43% (13/30), a false positive rate of 57% (17/30), and a false negative rate of 26% (5/19).

2.3.3 Gettings et al. (1996)

Based on the reclassification process, 24 of the 25 substances tested in this study could be assigned a EU classification (**Table I-4**). The remaining substance had inadequate *in vivo* data for assigning a classification according to the EU system (EU [2001]). For the 24 substances that could be evaluated, the IRE test method has an accuracy of 67% (16/24), a sensitivity of 63% (10/16), a specificity of 75% (6/8), a false positive rate of 25% (2/8), and a false negative rate of 38% (6/16).

2.3.4 Guerriero et al. (2004)

The original IRE test method accuracy analysis included 44 substances. Upon reclassification, sufficient data were available to permit EU classification on 38 of the 44 original substances and were used for the accuracy analysis (**Table I-4**). The remaining six substances had inadequate *in vivo* data for assigning a classification according to the EU system (EU [2001]). For the 38 substances, the IRE test method has an accuracy of 79% (30/38), sensitivity of 100% (11/11), specificity of 70% (19/27), a false positive rate of 30% (8/27), and false negative rate of 0% (0/11).

2.3.5 Expanded Data Set

Subsequent to the original IRE test method accuracy analysis, the total data base of 149 substances was mined to established an expanded data set that included: (1) all substances evaluated by Guerriero et al. (2004) that could be assigned an EU classification (EU [2001]), and (ii) any additional substances classified as severe irritants by CEC (1991), Balls et al. (1995), and Gettings et al. (1996), based either on an *in vitro* corneal opacity score of at least 3.0 or an *in vitro* corneal swelling of at least 25%, that had corresponding *in vivo* rabbit eye test data that would allow the substances to be classified according to the EU system (EU [2001]). As noted previously, these two criteria were among those used by Guerriero et al. (2004) to identify corrosive/severe irritants. Rules for classifying a substance that was evaluated in multiple laboratories are the same as described in **Section I-2.1.4**.

Table I-4. Evaluation of the Performance of the IRE Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU[1] Classification System, by Study and Overall

Data Source	Data Set	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
CEC (1991)	New[4]	15/21	87	13/15	100	5/5	80	8/10	71	5/7	100	8/8	20	2/10	0	0/5
	Old[4]	21/21	86	18/21	100	8/8	77	10/13	73	8/11	100	10/10	23	3/13	0	0/8
Balls et al. (1995)[5,6]	New	49/59	55	27/49	74	14/19	43	13/30	45	14/31	72	13/18	57	17/30	26	5/19
	Old	59/59	53	31/59	67	14/21	45	17/38	40	14/35	71	17/24	55	21/38	33	7/21
Gettings (1996)	New	24/25	67	16/24	63	10/16	75	6/8	83	10/12	50	6/13	25	2/8	38	6/16
	Old	25/25	52	13/25	43	3/7	56	10/18	27	3/11	71	10/14	44	8/18	57	4/7
Guerriero et al. (2004)	New	38/44	79	30/38	100	11/11	70	19/27	58	11/19	100	19/19	30	8/27	0	0/11
	Old	44/44	77	34/44	100	15/15	66	19/29	60	15/25	100	19/19	34	10/29	0	0/15
Expanded Data Set[7]	New	80/91	70	56/80	100	37/37	44	19/43	61	37/61	100	19/19	56	24/43	0	0/37

[1]EU = European Union (EU [2001]).
[2]N = Number of substances included in this analysis/the total number of substances in the study.
[3]Data used to calculate the percentage.
[4]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft IRE BRD.
[5]One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice within the same laboratory. The results were discordant with respect to EU classification; the analysis was performed assuming an R41 classification.
[6]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.
[7]Includes the 38 substances tested by Guerriero et al. (2004) that could be classified and 42 additional substances classified as severe irritants from Balls et al. (1995) and Gettings et al. (1996), based either on an *in vitro* corneal opacity score of at least 3.0 or an *in vitro* corneal swelling of at least 25%; these were among the criteria used by Guerriero et al. (2004) to identify corrosive/severe irritants. When the same substance was evaluated in multiple laboratories, the IRE ocular irritancy potential for each independent test result was determined. Subsequently, an overall IRE ocular irritancy classification was assigned for each substance based on the majority of ocular irritancy classification calls and this call was used in the analysis of IRE test method accuracy (approach described in **Section I-2.0**); this process reduced the total number of substances in the expanded data set to 80 for the EU classification system (EU 2001).

Using this approach, the total number of substances in the expanded data set was 80 for the EU classification system (EU [2001]). For these 80 substances (**Table I-4**), the IRE test method has an accuracy of 70% (56/80), a sensitivity of 100% (37/37), a specificity of 44% (19/43), a false positive rate of 56% (24/43), and a false negative rate of 0% (0/37).

2.4 Accuracy of the IRE Test Method for the GHS Ocular Hazard Classification System, by Chemical Class and Property of Interest - Reanalysis

In order to further evaluate discordant responses of the IRE test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., pesticides, surfactants, pH, physical form). Because the international community will soon adopt the GHS classification system for hazard labeling (UN [2003]), and considering that there were only modest differences in overall IRE test method accuracy among the three regulatory classification systems (i.e., EPA, EU, GHS), these sub-analyses are focused only on the GHS classification system, using the Expanded Data Set (**Table I-5**).

Limiting this evaluation to chemical classes with at least 5 substances, the chemical classes that had the highest rate of IRE test method overprediction according the GHS classification system (i.e., were false positives) were ketones (67%, [4/6]), esters (67%, [4/6]), and alcohols (60%, [6/10]).

Ten surfactants were evaluated (seven cationic and 3 nonionic). Overall, surfactants had a false positive rate of 50% (2/4) and a false negative rate of 0% (0/6). Cationic surfactants had a false positive rate of 100% (1/1) and a false negative rate of 0% (0/6).

With regard to physical form of the substances overpredicted by the IRE test method, liquids had a higher overprediction rate (83%, [19/23]) than solids (25%, [5/20]). There was insufficient data to analyze the effect of pH on overprediction. The false positive rates may be exaggerated by the lack of inclusion of additional true negative substances to those tested by Guerriero et al. (2004).

No substances were underpredicted (i.e., were false negatives) by the IRE test method (for the Expanded Data Set) according to the GHS classification system (see **Table I-5**). Thus, an analysis of underprediction based on chemical class, physical form, pH, or NICEATM GHS Category I subclassification was not possible.

2.5 Accuracy of the IRE Test Method for Identifying Ocular Corrosives and Severe Irritants – Summary of Reanalysis

As detailed in **Section I-1.0**, no additional IRE test method data was received after the Expert Panel meeting on January 11 and 12, 2005. However, as recommended by the Expert Panel, a reanalysis was conducted on an expanded data set that included (1) all substances evaluated by Guerriero et al. (2004) that could be assigned an GHS/EPA/EU classification based on *in*

Table I-5. False Negative and False Positive Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the GHS[1] Classification System (Analysis Based on the Expanded Data Set)

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	76	56	24/43	0	0/33
Chemical Class[6]					
Alcohol	11	60	6/10	0	0/1
Amide	5	0	0/3	0	0/2
Amine	9	60	3/5	0	0/4
Carboxylic acid	5	67	2/3	0	0/2
Ester	6	67	4/6	-	0/0
Ether	8	40	2/5	0	0/3
Formulation	12	100	2/2	0	0/10
Heterocycle	16	50	4/8	0	0/8
Ketone	6	67	4/6	-	0/0
Onium compound	9	33	1/3	0	0/6
Sulfur compound	7	20	1/5	0	0/2
Properties of Interest					
Liquid/Solution	43	83	19/23	0	0/20
Solid	33	25	5/20	0	0/13
Surfactant – Total	10	50	2/4	0	0/6
-nonionic	3	50	1/2	0	0/1
-anionic	-	-	-	-	-
-cationic	7	100	1/1	0	0/6
pH – Total[7]	0	-	-	-	-
- acidic (pH < 7.0)	0	-	-	-	-
- basic (pH > 7.0)	0	-	-	-	-
NICEATM GHS Category 1 Subgroup[8]	21	-	-	0	0/0
- Total	4	-	-	0	0/4
- 4 (CO=4 at any time)	3	-	-	0	0/3
- 3 (severity/persistence)	2	-	-	0	0/2
- 2 (severity)	*9*	-	-	0	0/9
- *2-4 combined*[9]	12	-	-	0	0/12
- 1 (persistence)					

[1]GHS = United Nations Globally Harmonized System (UN [2003]).
[2]N = number of substances.
[3]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[4]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[5]Data used to calculate the percentage.
[6]Chemical classes included in this table are represented by at least five substances tested in the IRE test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh). See **Appendix B**.
[7]Total number of GHS Category 1 substances for which pH information was available.
[8]Subgroups assigned based on the whether classification as a GHS Category 1 substance was based on severity and/or persistence. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including Corneal Opacity [CO]=4); 3: based on lesions that are both severe (not including CO=4) and persistent; 4: CO = 4 at any time.
[9]Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.

vivo rabbit eye test data, and (ii) any additional substances classified by IRE as severe irritants by CEC (1991), Balls et al. (1995), and Gettings et al. (1996) and that could also be assigned a GHS/EPA/EU classification based on *in vivo* rabbit eye test data. For the additional substances, a severe irritant classification was based either on an *in vitro* corneal opacity score of at least 3.0 or an *in vitro* corneal swelling of at least 25%. These two criteria were among the four used by Guerriero et al. (2004) to identify corrosive/severe irritants (the other endpoints used by Guerriero et al. (2004) included fluorescein penetration and epithelial integrity). Substances that were not classified as severe irritants in these IRE studies by CEC (1991), Balls et al. (1995) and Gettings et al. (1996) could not be used in the reanalysis, because an evaluation of any one of the parameters not evaluated in the respective studies could have resulted in the substance being classified as a corrosive or severe irritant. For example, in Gettings et al. (1996), only corneal swelling was measured. Substances that produced corneal swelling of at least 25% were included in the "Expanded Data Set" and used in the reanalysis. However, a substance that did not produce ≥ 25% corneal swelling might have produced a corneal opacity score, fluorescein penetration score, or damage of the epithelium that would have classified it as a severe irritant had any of these endpoints been evaluated. Accordingly, because substances classified as nonsevere irritants in Gettings et al. (1996) could potentially be classified as severe irritants using these other criteria, such substances are not included in the Expanded Data Set analysis.

The reanalysis of the accuracy of the IRE test method for identifying ocular corrosives and severe irritants also took into account the reclassification of some nonsevere irritants as severe irritants (see **Section I-1.0** and **Appendix I-A**). As the changes in accuracy are independent of the ocular hazard classification system used, this discussion is limited to the GHS classification system.

When the reanalysis is restricted to Guerriero et al. (2004), the IRE test method version that evaluated the greatest number of endpoints, the reclassification changed from 78% (28/36) in the draft IRE BRD to 79% (30/38) in the reanalysis. The false negative rate stayed the same. (draft IRE BRD = 0% [0/12]; reanalysis: 0% [0/11]). The false positive rate decreased from 32% (8/24) in the draft IRE BRD to 30% (8/27) in the reanalysis.

With the addition of some substances classified as corrosive/severe irritants in Balls et al. (1995) and Gettings et al. (1996), the overall accuracy was 68% (52/76), the false negative rate was 0% (0/33), while the false positive rate was 56% (24/43) (i.e., the additional data included 38 substances classified by IRE as severe irritants, 22 of which were also severe irritants *in vivo* and 16 of which were nonsevere irritants or nonirritants *in vivo*). The expanded data set is potentially confounded by the exclusion of substances with true negative outcomes (matching *in vivo* and *in vitro* nonsevere or nonirritant classifications), which would affect both specificity and the false positive rate.

Table I-6 provides a summary of the revised analysis of the overall performance of the Expanded Data Set, when compared to the GHS classification system (UN [2003]). As noted from this analysis, the false positive substances included 11 nonirritants, three Category 2B substances, and 10 Category 2A substances. No severe irritants (0/33) were underpredicted.

Table I-6. Overall Accuracy of the IRE Test Method in Predicting the Irritancy of a Substance as Defined by the GHS[1] Classification System (Analysis Based on the Expanded Data Set)[2]

		In Vitro Classification		
		Severe Irritant	**Nonsevere Irritant**	**TOTAL**
***In Vivo* Classification[3]**	**1**	33	0	33
	2A	10	1	11
	2B	3	3	6
	Nonirritant	11	15	26
	TOTAL	57	19	76

[1]GHS = United Nations Globally Harmonized System (UN [2003].

[2]Includes the 38 substances tested by Guerriero et al. (2004) that could be classified and 38 additional substances classified as severe irritants from Balls et al. (1995) and Gettings et al. (1996), based either on an *in vitro* corneal opacity score of at least 3.0 or an *in vitro* corneal swelling of at least 25%; these were among the criteria used by Guerriero et al. (2004) to identify corrosive/severe irritants. When the same substance was evaluated in multiple laboratories, the IRE ocular irritancy potential for each independent test result was determined. Subsequently, an overall IRE ocular irritancy classification was assigned for each substance based on the majority of ocular irritancy classification calls and this call was used in the analysis of IRE test method accuracy (approach described in **Section I-2.0**); this process reduced the total number of substances in the expanded data set to 76 for the GHS classification system (UN [2003]).

[3]Thirty-four substances included in **Appendix I-A** had insufficient data with which to assign a GHS classification and therefore were not included in this table.

3.0 RELIABILITY OF THE IRE TEST METHOD - REANALYSIS

As discussed in the draft IRE BRD, an assessment of test method reliability (intralaboratory repeatability and intra- and inter-laboratory reproducibility) is an essential element of any evaluation of the performance of an alternative test method (ICCVAM [2003]). Repeatability refers to the closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period (ICCVAM [1997, 2003]). Intralaboratory reproducibility refers to the determination of the extent to which qualified personnel within the same laboratory can replicate results using a specific test protocol at different times. Interlaboratory reproducibility refers to the determination of the extent to which different laboratories can replicate results using the same protocol and test chemicals, and indicates the extent to which a test method can be transferred successfully among laboratories. A reliability assessment includes reviewing the rationale for selecting the substances used to evaluate test method reliability, a discussion of the extent to which the substances tested represent the range of possible test outcomes and the properties of the various substances for which the test method is proposed for use, and a quantitative and/or qualitative analysis of repeatability and intra- and inter-laboratory reproducibility. In addition, measures of central tendency and variation are summarized for historical control data (negative, vehicle, positive), where applicable.

3.1 Substances Used to Re-evaluate the Reliability of the IRE Test Method

An evaluation of the intralaboratory repeatability and reproducibility of the IRE test method could not be conducted in the original reliability analysis due to the lack of appropriate data (see draft IRE BRD, Nov 1, 2004). No additional IRE test method data was submitted in response to the *FR* notice (see **Section I-1.0**). However, due to the *in vivo* reclassification of some substances from nonsevere irritants/nonirritants to severe irritants and to the development of the Expanded Data Set (see **Section I-1.0**), a reanalysis of the reproducibility of the IRE test method was conducted. The sources of data available for conducting an assessment of IRE test method interlaboratory reproducibility were the EC/HO validation study from Balls et al. (1995) and the CEC (1991) prevalidation study. In the Balls et al. (1995) validation study, four laboratories evaluated the accuracy and reliability of the IRE test method using 60 substances (i.e., there were 52 different substances with four substances tested at two different concentrations and two substances tested at three different concentrations, for a total of 60 possible ocular irritation outcomes). One substance (thiourea) was tested *in vitro* in the IRE assay but, due to its excessive toxicity *in vivo*, was excluded from the comparison of *in vitro* and *in vivo* test results. In the CEC (1991) collaborative study, three laboratories evaluated the accuracy and reliability of the IRE test method using 21 substances.[8]

3.2 Reanalysis of IRE Test Method Intralaboratory Repeatability

Generally, analyses of intralaboratory repeatability have included approaches such as:

- a coefficient of variation (CV) analysis, which is a statistical measure of the deviation of a variable from its mean (e.g., Holzhütter et al. [1996])
- analysis of variance (ANOVA) methods (e.g., Holzhütter et al. [1996]; ASTM [1999]).

Due to the lack of available IRE test data for replicate enucleated rabbit eyes within individual experiments performed by the same laboratory and for repeat experiments conducted on the same substance under exactly the same conditions, an evaluation of the intralaboratory repeatability of the IRE test method could not previously be conducted (see draft IRE BRD). As noted above, no additional data were received that would enable an analysis of intralaboratory repeatability.

3.3 Reanalysis of IRE Test Method Intralaboratory Reproducibility

Generally, analyses of intralaboratory reproducibility have included approaches such as:

- CV analysis, which is a statistical measure of the deviation of a variable from its mean (e.g., Holzhütter et al. [1996])
- ANOVA methods (e.g., Holzhütter et al. [1996]; ASTM [1999]).

[8] Some severely irritating substances tested by the CEC (1991) were excluded from this evaluation due to the lack of individual *in vivo* rabbit eye data. Classification of these substances had been based on ocular effects in humans, dermal studies, or pH.

Due to the lack of available IRE test data for experiments conducted multiple times in the same laboratory, an evaluation of IRE test method intralaboratory reproducibility could not conducted in the original IRE BRD (see draft IRE BRD). No additional IRE data has been received that would enable an evaluation of intralaboratory reproducibility.

3.4 Reanalysis of IRE Test Method Interlaboratory Reproducibility

Generally, analyses of interlaboratory variability have included approaches such as:

- the extent of concordance among laboratories in assigning the same regulatory classification for a particular substance (e.g., Holzhütter et al. [1996])
- a CV analysis, which is a statistical measure of the deviation of a variable from its mean (e.g., Holzhütter et al. [1996])
- ANOVA methods (e.g., Holzhütter et al. [1996]; ASTM [1999])
- bivariant scatter diagrams/correlation analyses for pairs of laboratories to assess the extent possibility of divergence (e.g., Holzhütter et al. 1996)

3.4.1 Qualitative Assessment of Interlaboratory Reproducibility

Data from CEC (1991) and Balls et al. (1995) were used to qualitatively evaluate the interlaboratory reproducibility of the IRE test method. For an assessment of interlaboratory reproducibility, substances classified as corrosive/severe irritants or nonsevere irritants/nonirritants were further classified within the EPA, EU, and GHS classification systems (EPA [1996]; EU [2001]; UN [2003]) by their *in vivo* rabbit eye test results. Because the focus of this assessment is on the interlaboratory reproducibility of the IRE test method in identifying corrosives/severe irritants versus nonsevere irritants/ nonirritants, considerable variability could exist among laboratories in their classification of substances as nonsevere irritants or nonirritants (e.g., three laboratories could classify a substance as a nonirritant and one laboratory could classify the same substance as a moderate irritant; for the purpose of the analysis, this would be considered 100% agreement between laboratories).

3.4.1.1 *GHS Ocular Hazard Classification System*

For this classification system, one study could be used to assess the interlaboratory reproducibility of the IRE test method: Balls et al. (1995). The four participating laboratories in the EC/HO study (Balls et al. [1995]) were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 35 (59%) of the 59 substances tested (see **Table I-7**).

As shown in **Table I-7:**

- All four participating laboratories agreed on the classification of 14 (100%) of the 14 substances that were GHS corrosives/severe irritants[9].
- Five (55%) of the nine substances classified according to the GHS based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by all four participating laboratories as nonsevere irritants (i.e., Category 2A and

[9] The overall *in vitro* classification for each substance was determined based on the most frequent individual laboratory classification, or in the case of an even number of discordant responses, the most severe classification.

2B irritants) or nonirritants, whereas four of the nine substances (44%) had 75% agreement among the laboratories. The five substances incorrectly classified by all four laboratories were Captan 90 concentrate, dibenzoyl-L-tartaric acid, 2,5-dimethylhexanediol, 15% sodium lauryl sulfate, and sodium perborate tetrahydrate.

- Eight (40%) of the 20 substances classified according to the GHS based on *in vivo* rabbit eye data as nonsevere irritants were incorrectly classified by the four laboratories as corrosives or severe irritants. Of the 12 substances (60%) with discordant results among the four laboratories, three (15%) (ethyl acetate, iso-propanol, and methyl acetate) were incorrectly classified by three of the four laboratories and nine (45%) (acetone, 0.1% cetylpyridinium bromide, ethyl-2-methylacetoacetate, Fomesafen, Maneb, methylisobutylketone, n-octanol, polyethylene glycol 400, and toluene) were incorrectly classified by two of the four laboratories.
- All four laboratories agreed on the classification of six (43%) of the 14 substances classified as GHS nonsevere irritants/nonirritants. Of the eight substances (57%) with discordant classification results, all eight substances (ammonium nitrate, butyl acetate, dibenzyl phosphate, 2,6-dichorobenzoyl chloride, methyl acetate, tetra-aminopyrimidine sulfate, 3% trichloroacetic acid, and Tween 20) were correctly classified by three of the four laboratories.
- Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), two (3%) of the 59 test substances could not be classified according to the GHS classification scheme. All four laboratories were in agreement with the classification of one of these substances as nonsevere irritant/nonirritant and one substance as a corrosive/severe irritant.

3.4.1.2 *EPA Ocular Hazard Classification System*

The four participating laboratories in the EC/HO study (Balls et al. [1995]) were in 100% agreement for the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 36 (61%) of the 59 substances tested. As shown in **Table I-8**:

- All four participating laboratories agreed on the classification of 18 (100%) of the 18 substances that were EPA corrosives/severe irritants[10].
- Four (57%) of the seven substances classified according to the EPA (1996) based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by all four participating laboratories as nonsevere irritants/nonirritants. Three substances (43%) were shown to have discordant *in vitro* classification results among the four participating laboratories (Captan 90 concentrate, 2,5-dimethylhexanediol, and sodium lauryl sulfate [15%]). These substances were incorrectly classified by three of the four laboratories.

[10] As described in **Section I-2.0**, the overall *in vitro* classification for each substance was determined based on the most frequent individual laboratory classification, or in the case of an even number of discordant responses, the most severe classification.

Table I-7. Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the GHS[1] Classification System

Classification (*in vivo*/ *in vitro*)[2]	Data Set	Number of Substances	Number of Testing Labs	Substances with 100% Agreement Among Labs	Substances with 75% Agreement Among Labs	Substances with 50% Agreement Among Labs
+/+	New[3]	14	4	14 (100%)	0 (0%)	0 (0%)
	Old[3]	14	4	14 (100%)	0 (0%)	0 (0%)
+/-	New	9	4	5 (55%)	4 (44%)	0 (0%)
	Old	8	4	4 (50%)	4 (50%)	0 (0%)
-/+	New	20	4	8 (40%)	3 (15%)	9 (45%)
	Old	20	4	8 (40%)	3 (15%)	9 (45%)
-/-	New	14	4	6 (43%)	8 (57%)	0 (0%)
	Old	14	4	6 (43%)	8 (57%)	0 (0%)
?/-	New	1	4	1 (100%)	0 (0%)	0 (0%)
	Old	2	4	2 (100%)	0 (0%)	0 (0%)
?/+	New	1	4	1 (100%)	0 (0%)	0 (0%)
	Old	1	4	1 (100%)	0 (0%)	0 (0%)
TOTAL	New	59	4	35 (59%)	15 (25%)	9 (15%)
	Old	59	4	35 (59%)	15 (25%)	9 (15%)

[1]GHS = Globally Harmonized System (UN [2003]).

[2]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category 1); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category 2A, 2B) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), a GHS classification could not be made. See **Section 2.0** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[3]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft IRE BRD.

- Eight (40%) of the 20 substances classified according to the EPA based on *in vivo* rabbit eye data as a nonsevere irritant/nonirritant were incorrectly classified by all four participating laboratories as a corrosive/severe irritant. Of the 12 remaining substances (60%), three substances ((15%) ethyl acetate, iso-propanol, and methyl acetate) were incorrectly classified by three of the four laboratories and nine substances (45%) (acetone, cetylpyridinium bromide, ethyl-2-methylacetoacetate, Fomesafen, Maneb, methylisobutylketone, n-octanol, polyethylene glycol 400, and toluene) by two of the four laboratories.
- Six (43%) of the 14 substances classified according to the EPA (1996) based on *in vivo* rabbit eye data as nonsevere irritants or nonirritants were correctly classified by all four laboratories. All eight substances (57%) with discordant classification results (ammonium nitrate, butyl acetate, dibenzyl phosphate, 2,6-dichlorobenzoyl chloride, methyl acetate, tetra-aminopyrimidine sulfate,

3% trichloroacetic acid, and Tween 20) were correctly classified by three of the four laboratories.

- The three substances classified *in vitro* as nonsevere irritants and the two substances classified as corrosives or severe irritants, which originally could not be assigned an *in vivo* classification, were reclassified as severe irritants and correctly identified by all four laboratories. These substances were 2,2-dimethylbutanoic acid, imidazole, promethazine, and pyridine.

Table I-8. Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the EPA[1] Classification System

Classification (*in vivo*/ *in vitro*)[2]	Data Set	Number of Substances	Number of Testing Labs	Substances with 100% Agreement Among Labs	Substances with 75% Agreement Among Labs	Substances with 50% Agreement Among Labs
+/+	New[3]	18	4	18 (100%)	0 (0%)	0 (0%)
	Old[3]	13	4	13(100%)	0 (0%)	0 (0%)
+/-	New	7	4	4 (57%)	3 (43%)	0 (0%)
	Old	7	4	4 (57%)	3 (43%)	0 (0%)
-/+	New	20	4	8 (40%)	3 (15%)	9 (45%)
	Old	20	4	8 (40%)	3 (15%)	9 (45%)
-/-	New	14	4	6 (43%)	8 (57%)	0 (0%)
	Old	14	4	6 (43%)	8 (57%)	0 (0%)
?/-	New	0	4	0 (0%)	0 (0%)	0 (0%)
	Old	3	4	2 (66%)	1 (33%)	0 (0%)
?/+	New	0	4	0 (0%)	0 (0%)	0 (0%)
	Old	2	4	2 (100%)	0 (0%)	0 (0%)
TOTAL	New	59	4	36 (61%)	14 (24%)	9 (15%)
	Old	59	4	35 (59%)	15 (25%)	9 (15%)

[1]EPA = U.S. Environmental Protection Agency (EPA [1996]).

[2]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category I); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category II, III) or nonirritant (category IV); a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), an EPA classification could not be made. See **Section I-2.0** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[3]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft IRE BRD.

3.4.1.3 *EU Ocular Hazard Classification System*

Using the Balls et al. (1995) data set, the participating laboratories were in 100% agreement with regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 37 (63%) of the 59 substances tested. As shown in **Table I-9**:

- All four participating laboratories agreed on the classification of 12 (100%) of the 12 substances that were EU corrosives/severe irritants

- Three (50%) of the six substances classified according to the EU based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by all four participating laboratories as nonsevere irritants/nonirritants. Of the three substances (50%) with discordant *in vitro* classification results among the four participating laboratories, all three substances (Captan-90 concentrate, dibenzoyl-L-tartaric acid, and 2,5-dimethylhexanediol) were incorrectly classified by three of the four laboratories.
- Seven (39%) of the 18 substances classified according to the EU based on *in vivo* rabbit eye data as a nonsevere irritants/nonirritant was incorrectly classified by all four participating laboratories as a corrosives/severe irritant. Of the 11 substances (61%) with discordant *in vitro* classification results among the four participating laboratories, two substances (44%), ethyl acetate and methyl acetate, were incorrectly classified by three laboratories and nine (50%) were incorrectly classified by two of the four laboratories (acetone, γ-butyrolactone, 0.1% cetylpyridinium bromide, ethyl-2-methylacetoacetate, Fomesafen, methylisobutylketone, n-octanol, polyethylene glycol 400, and toluene).
- All four laboratories agreed on the classification of six (50%) of the 12 substances classified as EU nonsevere irritants/nonirritants the four participating laboratories. Three of the four laboratories were in agreement for the six substances (50%) with discordant classification results (ammonium nitrate, 4-carboxybenzaldehyde, dibenzyl phosphate, tetra-aminopyrimidine sulfate, 3% trichloroacetic acid, and Tween 20).
- Four of six (67%) of substances classified *in vitro* as nonirritants, but could not be classified *in vivo* due to the lack of sufficient data, were classified as such by all four laboratories. Two of the six (33%) were classified as nonirritants *in vitro* by two of three laboratories.
- Five of five (100%) of substances were classified *in vitro* as corrosives or severe irritants by all four laboratories, but could not be classified *in vivo* due to the lack of appropriate data.

Table I-9. Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the EU[1] Classification System

Classification (*in vivo*/ *in vitro*)[2]	Data Set	Number of Substances	Number of Testing Labs	Substances with 100% Agreement Among Labs	Substances with 75% Agreement Among Labs	Substances with 50% Agreement Among Labs
+/+	New[3]	12	4	12 (100%)	0 (0%)	0 (0%)
	Old[3]	14	4	14 (100%)	0 (0%)	0 (0%)
+/-	New	6	4	3 (50%)	3 (50%	0 (0%)
	Old	7	4	4 (57%)	3 (43%)	0 (0%)
-/+	New	18	4	7 (39%)	2 (44%)	9 (50%)
	Old	21	4	9 (43%)	3 (14%)	9 (43%)
-/-	New	12	4	6 (50%)	6 (50%)	0 (0%)
	Old	17	4	9 (53%)	8 (47%)	0 (0%)
?/-	New	6	4	4 (67%)	2 (33%)	0 (0%)
	Old	0	4	0 (0%)	0 (0%)	0 (0%)
?/+	New	5	4	5 (100%)	0 (0%)	0 (0%)
	Old	0	4	0 (0%)	0 (0%)	0 (0%)
TOTAL	New	59	4	37 (63%)	13 (22%)	9 (15%)
	Old	59	4	36 (61%)	14 (24%)	9 (15%)

[1]EU = European Union (EU [2001]).
[2]A "+" indicates that the substance was assigned an overall classification of corrosive or severe irritant (Category R41); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category R36) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data, an EU classification could not be made. See **Section I-2.0** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.
[3]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft IRE BRD.

Using the CEC (1991) data set, the three participating laboratories were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 17 (81%) of the 21 substances tested (**Table I-10**).

As shown in **Table I-10**:

- Three (60%) of five substances were classified according to *in vivo* rabbit eye data as corrosives or severe irritants and these were identified correctly by all three laboratories. One discordant substance (sodium dodecyl sulfate) was correctly classified by two of the three laboratories, and one (dibutyltin chloride) was correctly classified by one of two laboratories.
- No substances were identified as false positives (i.e., as severe irritants *in vivo* and as nonsevere irritants *in vitro*).
- Two of two (100%) nonsevere irritants *in vivo* were incorrectly classified as severe irritants *in vitro* by all three laboratories. There were no discordant substances.

- Six of eight (75%) substances were in complete agreement among laboratories for identification of nonsevere irritants. Two discordant substances (25%) (Brij-35 and 2-butoxyethylacetate) were identified as nonsevere irritants by two of the three testing laboratories.
- All three laboratories agreed in the identification of two substances as nonsevere irritants (100%) and another four as severe irritants, although no *in vivo* classification could be assigned to these substances.

Table I-10. Interlaboratory Variability of CEC Collaborative Study (1991) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the EU[1] Classification System

Classification (*in vivo*/ *in vitro*)[2]	Data Set	Number of Substances	Number of Testing Labs	Substances with 100% Agreement Among Labs	Substances with 67% Agreement Among Labs	Substances with 33% Agreement Among Labs
+/+	New[3]	5	3	3 (60%)	1 (20%)	1 (20%)[4]
	Old[3]	8	3	7 (88%)	1 (12%)	0 (0%)
+/-	New	0	3	0 (0%)	0 (0%)	0 (0%)
	Old	0	3	0 (0%)	0 (0%)	0 (0%)
-/+	New	2	3	2 (100%)	0 (0%)	0 (0%)
	Old	3	3	3 (100%)	0 (0%)	0 (0%)
-/-	New	8	3	6 (75%)	2 (25%)	0 (0%)
	Old	10	3	8 (80%)	2 (20%)	0 (0%)
?/-	New	2	2[5]	2 (100%)	0 (0%)	0 (0%)
	Old	0	-	0 (0%)	0 (0%)	0 (0%)
?/+	New	4	3	4 (100%)[6]	0 (0%)	0 (0%)
	Old	0	-	0 (0%)	0 (0%)	0 (0%)
TOTAL	New	21	3	17 (81%)	3 (14%)	1 (5%)
	Old	21	3	18 (86%)	3 (14%)	0 (0%)

[1]EU = European Union (EU [2001]).
[2]A "+" indicates that the substance was assigned an overall classification of corrosive or severe irritant (Category R41); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category R36) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data, an EU classification could not be made. See **Section 2.0** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.
[3]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft IRE BRD.
[4]Agreement was among one of two laboratories (50% not 33%). The third laboratory did not test the material.
[5]Two of the three testing laboratories evaluated these two substances.
[6]One of the four substances was tested in two laboratories with severe classification assigned.

3.4.2 Quantitative Assessment of Interlaboratory Reproducibility

As detailed in the draft IRE BRD, to provide a quantitative assessment of interlaboratory variability, individual laboratory IRE test results were used to calculate a mean, standard deviation, and the %CV for corneal opacity, fluorescein retention, corneal swelling, and the

irritation index for each of the 59 substances tested in the Balls et al. (1995) study. Mean and median %CV values were calculated to provide an assessment of overall variability. This analysis was not affected by the information received subsequent to the release of the draft IRE BRD, and therefore is not presented here.

3.4.3 Additional Reanalyses of Interlaboratory Reproducibility

A comparison of the corneal opacity and corneal swelling measurements at one and four hours for substances that were tested in both the Balls et al. (1995) and Guerriero et al. (2004) data sets is presented in **Table I-11**. Correlation coefficients for corneal opacity scores at 1 and 4 hours were 0.77 and 0.78, respectively. Correlation coefficients for corneal swelling at 1 and 4 hours were 0.92 and 0.68, respectively. The corneal swelling measurements in Balls et al. (1995) were more variable than those in Guerriero et al. (2004). This might be attributed to differences in the methods of measurement of corneal thickness in the four contributing laboratories in the Balls et al. (1995) study employed amongst the various laboratories in this study to quantify corneal swelling (i.e., ultrasonic pachymeter vs. depth measuring gauge).

The draft IRE BRD also contains a description of the analysis performed by Balls et al. (1995) in which they determined the interlaboratory correlation between IRE test method endpoint data generated by each laboratory for all substances tested, as well as for subsets of test substances (water-soluble, water-insoluble, surfactants, solids, solutions, and liquids). This analysis was not affected by the information received subsequent to the release of the draft IRE BRD and therefore is not presented here.

3.5 IRE Test Method Historical Positive and Negative Control Data - Reanalysis

Concurrent positive control substances have not been employed in the IRE test method, and therefore, an evaluation of historical positive control data is not possible. One eye is traditionally included in each study as a negative/vehicle controls (isotonic saline).

3.6 Reliability of the IRE Test Method for Identifying Ocular Corrosives and Severe Irritants – Summary of Reanalysis

In the draft IRE BRD, no data was provided for the assessment of intralaboratory repeatability and reproducibility. Since no additional data was submitted for the IRE test method following the Expert Panel meeting, additional analyses of intralaboratory reliability could not be conducted.

The original IRE test method reliability analysis included an evaluation of interlaboratory reproducibility using both qualitative and quantitative approaches. While the quantitative analysis was unaffected by the reclassification of some nonsevere irritants/nonirritants as severe irritants, the qualitative analysis (correct classification as an ocular corrosive/severe irritant or as a non-corrosive/nonsevere irritant) of the individual laboratory test results obtained for the EC/HO validation study (Balls et al. [1995]) and for the CEC (1991) collaborative study were affected. Overall, in the Balls et al. (1995) study, the number of substances with 100% agreement between the four laboratories was 59-61% (35-36/59) in the

Table I-11. Interlaboratory Reproducibility of Corneal Endpoint Measures for Substances Tested in Common Between IRE Test Method Studies

***In Vitro* IRE Data**					
Test Material	**Endpoint**	**Balls et al. (1995)[1]**		**Guerriero et al. (2004)[2]**	
		Mean ± SD[3]			
		1 Hour	**4 Hours**	**1 Hour**	**4 Hours**
Sodium Hydroxide (10%)	Corneal Opacity Score[4]	3 ± 1.7	4 ± 0.0	3 ± 0.6	3 ± 0.6
	Swelling (%)[5]	102 ± 13.6	138 ± 25.3	111 ±28.8	NT[6]
Trichloroacetic Acid (30%)	Corneal Opacity Score	2 ± 0.5	2 ± 0.5	4 ± 0.0	4 ± 0.0
	Swelling (%)	24 ± 28.3	44 ± 33.1	12 ±1.3	54 ±9.5
Acetone	Corneal Opacity Score	0 ± 0.5	1 ± 1.1	2 ± 0.6	2 ± 0.6
	Swelling (%)	15 ± 12.0	32 ± 30.2	19 ± 9.7	50 ± 44.8
Allyl Alcohol[7]	Corneal Opacity Score	1 ± 0.9	2 ± 1.0	3 ± 0.6	3 ± 0.0
	Swelling (%)	16 ± 11.2	36 ± 20.2	41 ± 3.7	77 ± 2.9
n-Butanol	Corneal Opacity Score	1 ± 0.9	3 ± 0.6	2 ± 0.6	3 ± 0.6
	Swelling (%)	25 ± 11.0	75 ± 19.6	55 ± 5.9	92 ±19
Ammonium Nitrate	Corneal Opacity Score	0 ± 0.0	0 ± 0.0	0 ± 0.0	0 ± 0.0
	Swelling (%)	7 ± 3.1	10 ± 11.3	11 ± 1.4	15 ±3.4
Cetylpyridinium Bromide (10%)	Corneal Opacity Score	1 ± 0.9	2 ± 0.8	1 ± 0.6	1 ± 0.6
	Swelling (%)	18 ± 6.4	43 ± 29.4	49 (n=1)[8]	31 (n=1)
Methyl Ethyl Ketone	Corneal Opacity Score	1 ± 0.8	2 ± 0.4	3 ± 0.0	3 ± 0.0
	Swelling (%)	21 ± 6.3	61 ± 20.7	35 ± 8.3	105 ± 18.6

***In Vitro* IRE Data**					
Test Material	**Endpoint**	**Balls et al. (1995)[1]**		**Guerriero et al. (2004)[2]**	
		Mean ± SD[3]			
		1 Hour	**4 Hours**	**1 Hour**	**4 Hours**
Butyl Acetate	Corneal Opacity Score	0 ± 0.0	0 ± 0.4	0 ± 0.0	1 ± 0.6
	Swelling (%)	7 ± 4.9	15 ± 10.9	20 ± 1.3	30 ± 2.3
Toluene	Corneal Opacity Score	0 ± 0.51	0 ± 0.6	0 ± 0.0	0 ± 0.0
	Swelling (%)	14 ± 9.4	23 ± 13.9	7.4 ± 1.5	15 ± 2.6
Glycerol	Corneal Opacity Score	0 ± 0.0	0 ± 0.47	0 ± 0.0	0 ± 0.0
	Swelling (%)	8 ± 12.1	8 ± 14.7	13 ± 5.1	21 ± 4.6
Polyethylene Glycol 400	Corneal Opacity Score	0 ± 0.5	1 ± 0.6	0 ± 0.0	0 ± 0.0
	Swelling (%)	15 ± 12.1	18 ± 14.7	10 ± 1.9	16 ± 2.4

[1]Data were provided as mean scores of three isolated rabbit eyes from each of four laboratories. The mean corneal opacity score and corneal swelling measurement (and its standard deviation) of the four laboratories were then calculated.
[2]Data were provided as mean scores of three isolated rabbit eyes from which the standard deviation was calculated.
[3]SD = Standard deviation.
[4]Corneal opacity score represents a scale of 1-4.
[5]Corneal swelling was measured by either ultrasonic pachymeter or by depth gauge measurements in the Balls et al. (1995) study and by ultrasonic pachymeter in the Guerriero et al. (2004) study.
[6]NT = Not tested.
[7]Allyl alcohol was not used in the accuracy or reliability analyses because rabbit data from the Guerriero et al. (2004) study was not available.
[8]n = Number of eyes tested.

original analysis and 59-63% (35-37/59) in the reanalysis. The number of substances with 75% agreement between laboratories was 24-25% (14-15/59) in the original analysis and to 22-25% (13-15/59) in the reanalysis. The number of substances with 50% agreement between four laboratories did not change due to the reanalysis (15% [9/59 substances]).

Overall, in the CEC (1991) study, the number of substances with 100% agreement among the three laboratories decreased from 86% (18/21) to 81% (17/21) in the reanalysis. The number of substances with 67% agreement among the three laboratories remained the same at 14% (3/21), while the number of substances with 33% agreement was increased from 0% to 5% (1/21).

4.0 REFERENCES

ASTM. 1999. Standard practice for conducting an interlaboratory study to determine the precision of a test methods. ASTM E691-92. In: Annual Book of ASTM Standards. Philadelphia, PA: American Society for Testing and Materials.

Balls M, Botham PA, Bruner LH, Spielmann H. 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test. Toxicol In Vitro 9:871-929.

CEC. 1991. Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test. Doc. XI/632/91/V/E/1/131/91 Part I and II.

EPA. 1996. Label Review Manual: 2nd Ed. EPA737-B-96-001. Washington, DC: U.S. Environmental Protection Agency.

EU. 2001. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances. Official Journal of the European Communities L255:1-333.

Gettings SD, Lordo RA, Hintze KL, Bagley DM, Casterton PL, Chudkowski M, Curren RD, Demetrulias JL, Dipasquale LC, Earl LK, Feder PI, Galli CL, Glaza SM, Gordon VC, Janus J, Kurtz PJ, Marenus KD, Moral J, Pape WJW, Renskers KJ, Rheins LA, Roddy MT, Rozen MG, Tedeschi JP, Zyracki J. 1996. The CTFA evaluation of alternatives program: An evaluation of *in vivo* alternatives to the Draize primary rabbit eye irritation test. (Phase III) Surfactant-based formulations. Food Chem Toxicol 34:79-117.

Guerriero FJ, Seaman CW, Olson MJ, Guest R, Whittingham A. 2004. Retrospective assessment of the rabbit enucleated eye test (REET) as a screen to refine worker safety studies. [Abstract No. 1282] Toxicol Sci (The Toxicologist Supplement) 78(1-S).

Holzhütter HG, Archer G, Dami N, Lovell DP, Saltelli A, Sjöström M. 1996. Recommendations of the application of biostatistical methods during the development and

validation of alternative toxicological methods. ECVAM Biostatistics Task Force Report 1. ATLA 24:511-530.

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No: 03-4508. Research Triangle Park, NC: National Toxicology Program. Available: http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm [accessed 12 July 2005].

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Methods: A Report of the *ad hoc* Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No: 97-3981. Research Triangle Park, NC: National Toxicology Program. Available: http://iccvam.niehs.nih.gov/docs/guidelines/validate/pdf [accessed 12 July 2005].

MeSH Medical Subject Heading. Available: http://www.nlm.nih.gov/mesh [accessed 12 July 2005].

NICEATM. 2004. Draft Background Review Document. Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: The Isolated Rabbit Eye (IRE) Test Method. Available: http://iccvam.niehs.nih.giv/methods/ocudocs/ocu_brd.htm [accessed 12 July 2005].

NIEHS. 2005. Second request for data on chemicals evaluated by *in vitro* or *in vivo* ocular irritancy test methods. Fed Reg 70:9661-9662. Available: http://iccvam.niehs.nih.gov/methods/eyeirrit.htm [accessed 12 July 2005].

NIEHS. 2004. Request for public comment on the nomination for ocular toxicity test methods and related activities and request for data on chemicals evaluated by *in vitro* or *in vivo* ocular irritancy test methods. Fed Reg 69:13859-13861. Available: http://iccvam.niehs.nih.gov/methods/eyeirrit.htm [accessed 12 July 2005].

UN. 2003. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). New York & Geneva: United Nations Publications. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_ehtml. [accessed 12 July 2005].

[This Page Intentionally Left Blank]